

From: [Aaron Z. Tobin](#)
To: NelsonC@gtlaw.com; billk@kirkmanlawfirm.com
Cc: lars.berg@kellyhart.com; dee.kelly@kellyhart.com
Subject: RE: AA / Skiplagged / Data [IWOV-Interwoven.FID461762]
Date: Monday, July 1, 2024 12:20:23 PM
Attachments: [image001.png](#)

Dee and Cam – following up on my previous emails, and conversation with Dee Friday morning, my client has decided to voluntarily do what we offered American on Friday - upscale and enhance the production of the data. This process has resulted in enhanced OCR capabilities with very high accuracy rates (over 95%). My client started the process over the weekend, and we are producing this enhanced version in the following link <https://drive.google.com/> [REDACTED]

[REDACTED]. This production is without prejudice to Skiplagged's position that it fully complied with the Judge Ray's May 1 order.

We believe this should moot the complaints identified in the motion American filed on Friday evening. Please let me know as soon as possible if American agrees that the motion is now moot and when it will inform the court accordingly. Bill and I are available to discuss.

Thank you,

Aaron Z. Tobin
www.condontobin.com
214-265-3851

From: Aaron Z. Tobin
Sent: Friday, June 28, 2024 6:56 AM
To: NelsonC@gtlaw.com; billk@kirkmanlawfirm.com
Cc: lars.berg@kellyhart.com; dee.kelly@kellyhart.com
Subject: RE: AA / Skiplagged / Data [IWOV-Interwoven.FID461762]

Cam – after having a chance to hear Mr. Fuller's testimony today on his concerns about the quality of the images and OCR function on the data produced, we visited again internally. While Skiplagged maintains that it has complied with Judge Ray's May 1 Order, it does want to do everything it can within reason to try and reach a compromise to avoid further burdening the court with the issue. In that vein, my client is willing to take on the burden of writing additional code which will enhance/sharpen the text and thus reduce the noise in the existing production. With the font being a standard Times New Roman, this should substantially enhance the OCR function within the production and make it more effective. While no OCR function is perfect, this should increase the accuracy rate to 95% or better. We believe this process can be done in relatively short order. Is this offer acceptable to put to bed the dispute over the data produced and whether the production complies with the May 1 Order?

Aaron Z. Tobin
www.condontobin.com

214-265-3851

From: Aaron Z. Tobin

Sent: Wednesday, June 26, 2024 10:28 AM

To: NelsonC@gtlaw.com; billk@kirkmanlawfirm.com

Cc: lars.berg@kellyhart.com; dee.kelly@kellyhart.com

Subject: RE: AA / Skiplagged / Data [IWOV-Interwoven.FID461762]

Cam,

Thank you for your email. As we discussed, AA did not request information responsive to RFPs 4 and 5 and ROG 8 be produced in a "single, tab-delimited text file, with headers" until our 6/21 telephone call. We have considered AA's recent request and believe that it is untimely and outside Judge Ray's 5/1 Order, of which Skiplagged has fully complied with.

As you correctly quote, the May 1 Order required Skiplagged to produce the requested data "translated into a reasonably usable form, except to the extent that the translation requires analysis of the data." This is what Skiplagged did. Further, your email contains incorrect assumptions. First, you state that the "data exists in Skiplagged's systems in electronic, text format." This is not correct. The data exists in Skiplagged's systems, intertwined with other data, and in an encrypted form. Merely pulling the data would not be usable, instead it requires translation into a reasonably usable form. To do this, Mr. Zaman had to rent powerful computers with 360 processors, at a significant cost, to extract the relevant data and handle the encryption. The process of extracting and translating the Skiplagged data was done in many terabytes RAM using sophisticated data structures, not multiple steps as you assume. The python script Mr. Zaman wrote extracted, decrypted, and translated the data into the usable PDFs without creating an intermediate tab delimited format. No additional steps were taken to convert the data after it was extracted, decrypted and published to PDF. Indeed, as Mr. Zaman testified on June 12, 2024, [REDACTED]

As to the OCR issue raised, when copying and pasting from a PDF to a Word document, for example, it is not uncommon for there to be mistakes in the translation. Nothing was specifically done to interfere with the OCR. The OCR feature was implemented using one of the more accurate open-source OCR engines available that in turn made the data reasonably usable. We took a look at the sample you sent over and we were able to search for and find the PNR record locator as well as other information within the document using the find tool, a process that doesn't work without OCR. Further, fuzzy search can be done to overcome any OCR issues that you claim. And if AA is not satisfied with the OCR, it is my understanding that AA can run its own OCR process on the documentation produced.

In sum, AA very recently requested for the first time that Skiplagged produce the requested data in a particular format. This request comes well after the discovery requests, the motion to compel and May 1 Order. Skiplagged fully complied with the May 1 Order by producing the data requested in a reasonably usable format. To do so again, as AA is requesting would be unreasonably costly and unnecessarily burdensome.

Thank you,

Aaron Z. Tobin

www.condontobin.com

214-265-3851

From: NelsonC@gtlaw.com <NelsonC@gtlaw.com>
Sent: Friday, June 21, 2024 11:15 AM
To: billk@kirkmanlawfirm.com; Aaron Z. Tobin <atobin@condontobin.com>
Cc: lars.berg@kellyhart.com; dee.kelly@kellyhart.com
Subject: AA / Skiplagged / Data

Aaron,

This is in furtherance of our discussion today regarding the roughly 1.4 million PNR records that Skiplagged produced pursuant to the Court's order of May 1st. During our call, you asked me to put in an email exactly what American is asking for.

The short answer is (with more detailed explanation below), that American is simply asking for the data to be produced in a single, tab-delimited text file, with headers. To the extent their needs to be more than one row per PNR (for example, to show multiple segments) that would be acceptable.

The longer explanation is as follows.

On May 1st, the Court ordered that "Skiplagged also **SHALL PRODUCE** all data in its possession or control responsive to Interrogatory No. 8 and Request for Production Nos. 4-5 (except with regards to Skiplagged's customer's social security numbers or credit card numbers), translated into a reasonably useable form, except to the extent that the translation requires analysis of the data." In response to that order, Skiplagged produced 1.4 million PDFs.

These 1.4 million PDFs are not in a "reasonably useable form." Among other things, the OCR in the PDFs is inaccurate. I have attached just one example here. If you open the PDF, you can see that the PNR is "QSHWAU," but if you copy and paste this you get "QsHWwaU." There are many other OCR errors in this page alone, and there are similar errors throughout this production. These errors make it impossible for American to do any kind of analysis of the data.

What's more, we should not be talking about OCR in the first place. The data exists in Skiplagged's systems in electronic, text format. Converting that electronic data into images, and then applying OCR to it, were extra steps that never should have been carried out.

During the 30(b)(6) deposition of Skiplagged, Mr. Zaman [REDACTED]
[REDACTED]
[REDACTED]. He used a Python script to decrypt the encrypted fields. He could have stopped there, and simply sent us the decrypted, tab-delimited file.

Instead, Mr. Zaman inexplicably chose to take *extra* steps to convert the decrypted data into an image format, and then applying a sloppy OCR process to those images. Doing this rendered the data unreadable to a computer.

The data should have been produced in a single text file, in a format that would render it useable (i.e., so that it can be imported into Excel). Since Mr. Zaman has already admitted the data started in tab-delimited format, it should be produced that way.

Please give me a call if you have any questions. My mobile number is below.

Cam

Cameron M. Nelson
Shareholder

Greenberg Traurig, LLP
77 West Wacker Drive | Suite 3100 | Chicago, IL 60601
T +1 312.456.6590 | F +1 312.899.0360 | C +1 708.227.7106
NelsonC@gtlaw.com | www.gtlaw.com | [View GT Biography](#)



If you are not an intended recipient of confidential and privileged information in this email, please delete it, notify us immediately at postmaster@gtlaw.com, and do not use or disseminate the information.